Reaching and grasping with the tongue : Shared motor planning between hand actions and articulatory gestures

Vainio, Lari

2018-10

Vainio , L , Tiippana , K , Tiainen , M , Rantala , A & Vainio , M 2018 , ' Reaching and grasping with the tongue : Shared motor planning between hand actions and articulatory gestures ' , The Quarterly Journal of Experimental Psychology , vol. 71 , no. 10 , pp. 2129-2141 . https://doi.org/10.1177/1747021817738732

http://hdl.handle.net/10138/311362
https://doi.org/10.1177/1747021817738732

acceptedVersion

Reaching and grasping with the tongue: Shared motor planning between hand actions and articulatory gestures

Vainio, L., Tiippana, K., Tiainen, M., Rantala, A., & Vainio, M.

*Lari Vainio – the corresponding author; Email: lari.vainio@helsinki.fi; Tel: +358-2-94129392

*Kaisa Tiippana (kaisa.tiippana@helsinki.fi)

*Mikko Tiainen (mikko.o.tiainen@helsinki.fi)

*Aleksi Rantala (aleksi.rantala@helsinki.fi)

**Martti Vainio (martti.vainio@helsinki.fi)

* Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Haartmaninkatu 8, 00014, Finland

** Phonetics and speech synthesis research group, Department of Modern Languages, University of Helsinki, Unioninkatu 40, 00014, Finland

10.550 words

Abstract

Research has shown connections between articulatory mouth actions and manual actions. The present study investigates whether forward-backward hand movements could be associated with vowel production processes that program tongue fronting/backing, lip rounding/spreading, (Experiment 1) and/or consonant production processes that program tongue tip and tongue dorsum actions (Experiment 2). The participants had to perform either forward or backward hand movement and simultaneously pronounce different vowels or consonants. The results revealed a response benefit, measured in vocal and manual reaction times, when the responses consisted of front vowels and forward hand movements. Conversely, back vowels were associated with backward hand movements. Articulation of rounded vs. unrounded vowels or coronal vs. dorsal consonants did not produce the effect. In contrast, when the manual responses of forward-backward hand movements were replaced by precision and power grip responses, the coronal consonants [t] and [r] were associated with the precision grip while the dorsal consonant [k] was associated with the power grip. We propose that the movements of the tongue body, operating mainly for vowel production, share the directional action planning processes with the hand movements. Conversely, the tongue articulators related to tongue tip and dorsum movements, operating mainly for consonant production, share the action planning processes with the precision and power grip, respectively.

Keywords: speech, hand movements, articulation, language evolution, behavioral study

Introduction

The mouth-gesture theories suggest that there might be universal tendency to use outward pointing articulatory gestures, such as rounded vowels, in the words that semantically point away from one's own body (e.g., 'you') and inward pointing articulatory gestures, such as unrounded vowels, in the words that semantically point toward one's own body (e.g., 'me') (e.g., Wallace, 1881; Ramachandran & Hubbard, 2001). This view assumes that these kind of articulatory gestures mimic pointing hand gestures in speech, and might consequently be connected to processes that plan these hand gestures. The present study uses the dual-action paradigm to investigate whether this potential interaction can manifest itself in response advantage when participants have to produce a forward-backward hand movement and simultaneously pronounce a vowel. In particular, the study investigated whether forward-backward hand movements can be associated with articulatory processes that control frontness and/or roundness of vowel production. In addition, the study investigates whether the production of coronal consonants, produced by fronting the tip of the tongue, could be linked to forward hand movements and production of dorsal consonants, produced by raising the tongue body against the velum, could be linked to backward hand movements. Finally, the study also aims to provide further evidence for the recently proposed view (Vainio, Schulman, Tiippana, & Vainio, 2013; Tiainen, Tiippana, Vainio, Komeilipoor, & Vainio, 2017) that the processes related to the articulation of coronal consonants interact with the precision grip planning and dorsal consonants interact with the power grip planning.

Previously, the dual-action paradigms have been used to explore interactions between processes that program movements of two different body parts. For instance, there is tendency for moving, for example, finger, wrist and arm of two hands in the same direction rather than different directions (e.g., Swinnen, Jardin, Meulenbroek, Dounskaia & Hofkens-Van Den Brandt, 1997; Serrien, Bogaerts, Suy, & Swinnen, 1999). These kinds of movement symmetry effect might be based on representing locations of objects in space in terms of multi-sensory spatial codes that are involved in coordinated movements of multiple effectors such as hand, eyes and head (Cohen & Andersen, 2002). Regarding this spatial coding of movements, the most studied effectors are those involved in reaching (i.e., eye, head and hand) (Cohen & Andersen, 2002). However, articulation similarly requires spatially directed movements of oral organs (i.e., tongue, lips and jaw). For example, vowels can be identified according to the extent to which they are produced by moving the tongue within the horizontal axis (i.e., front vs. back vowels), moving the jaw and tongue within the vertical axis (i.e., close vs. open vowels), or by protruding or spreading the lips (i.e., rounded vs. unrounded vowels). In addition, the same posterior parietal (Colby & Goldberg, 1999) and premotor (Kawashima, Itoh, Ono, Satoh, Furumoto, Gotoh, et al., 1996) regions that are involved in spatial programming of hand movements have been also found to be involved in spatial programming of tongue movements (Corfield, Murphy, Josephs, Fink, Frackowiak, Guz et al., 1999; Watanabe, Sugiura, Miura, Watanabe, Maeda, Matsue, & Kawashima, 2004; Petrides & Pandya, 2009). That is, similar parieto-frontal mechanisms that are responsible for programming spatially directed hand movements might also be responsible for programming the spatial elements for articulatory gestures.

In addition, evidence suggests that there is an interaction between hand and mouth movements. For example, at the single-cell level, the same premotor neurons in monkey are involved in programming grasp actions performed with the hand and mouth (Rizzolatti, Camarda, Fogassi, Gentilucci, Luppino, & Matelli, 1988). Behavioural evidence additionally supports the view that there is an overlap in processes that program hand and mouth movements. For example, chimpanzees increasingly produce certain mouth movements such as protrusion and compression of the lips and tongue during fine manual manipulation (Waters & Fouts, 2002). In line with these findings, Higginbotham, Isaak, and Domingue (2008) have observed an increase in EMG responses of the orbicularis oris muscles –involved in producing articulations requiring lip protrusion– of the human participants during execution of the precision grasp. Moreover, Gentilucci and his colleagues have shown that when participants are required to grasp an object and simultaneously pronounce a meaningless syllable, the more the object requires finger opening for the grasp, the more the participants open their lips during vocalization (Gentilucci, Benuzzi, Gangitano, & Grimaldi, 2001).

Importantly, our recent evidence suggests that hand and mouth actions also interact at the level of programming movement direction (Vainio, Tiainen, Tiippana, Komeilipoor, & Vainio, 2015). In that study, the participants who were native Finnish speakers were required to pronounce, for example, the front-close vowel [i] or the back-open vowel [ɑ] and simultaneously perform a forward or backward hand movement with a joystick. We found that vocal and manual responses were performed relatively rapidly when the participant pronounced the front-close vowel [i] and made a forward hand movement. Conversely, vocal and manual responses were performed relatively rapidly when the participant pronounced the back-open vowel [ɑ] or the back-mid vowel [o] and made a backward hand movement. This articulation-movement effect was observed even though the participants were not required to explicitly process the frontness/backness dimension of the vowel, they were entirely unaware about the objectives of the study, and most of them did not even know that, for example, [i] is more a frontal vowel than [o]. Moreover, we have more recently observed the same effect in Czech speakers (Tiainen, Lukavský, Tiippana, Vainio, Šimko, Felisberti, & Vainio, 2017). These studies suggest that people have an implicit and language-independent tendency to link forward hand movements with front-close vowels and backward hand movements with back-open and back-mid-open vowels. That is, interaction in direction-related movement planning does not only operate in relation to two upper limbs (Swinnen et al., 1997; Serrien et al., 1999) or between hand-reach and saccadic eye movements (Cohen & Andersen, 2002) but there is similar interaction between hand and mouth actions.

The present study employs a dual-action choice reaction time (RT) task in order to explore further the interaction in planning directional articulatory movements and hand movements. As mentioned above, similar choice RT tasks have been commonly used to explore, for example, the movement symmetry effects in relation to two hands (Wakelin, 1976; Heuer, 1982). In general, these studies have been assumed to reveal how mechanisms related to separate actions couple in action planning (Rosenbaum, 1980). The basic principle behind this view is that when the task requires choosing between two alternative response options simultaneously for two different effectors, such as for the left and right hand, the action plans for all alternative responses are prepared prior to the stimulus onset. When the stimulus is presented, the prepared motor programs have to compete for the response execution. As a consequence, the time to select the required motor program for one action is affected by the response selection processes related to the other action. The more the motor programs required in a given trial share motor characteristics, the less there is interference between selection processes related to two actions, and consequently the faster both actions are executed. Thus, we assume that the dual-action that couples, for example, production of the front vowel and forward hand movement is performed faster than the dual-action that couples the production of the front vowel and backward hand movement because they share the same directional motor characteristics.

Hand movements and vowel production (Experiment 1)

So called mouth-gesture theories assume that speech has emerged to some extent in the process of the lips and tongue involuntarily mimicking one's own hand gestures (e.g., Paget, 1930; Hewes, 1973). According to our best knowledge, the oldest version of this view was proposed by Alfred Russel Wallace (1881). He noticed that "*many savages point with the lips as we do with the finger, signifying there, by protruding the lips in the direction to be indicated*" (p. 245). Similarly, Ramachandran and Hubbard (2001) suggested that articulatory gestures related to some words such as '*you*' might mimic outward pointing actions performed by a hand as the articulation of that word requires pouting one's lips outwards. In contrast, some words such as '*me*' might mimic inward pointing hand action as the articulation of that word requires spreading ones lips inwards. According to these views, a vocabulary might have been shaped, in the universal scale and over a long period of time, by these tendencies to mimic with articulators, for example, the outward-inward hand pointing actions.

We suggested in our previous study (Vainio et al., 2015) that these kinds of tendencies to mimic with articulators one's own deictic hand actions, might be based on the same neural interactions as the articulation-movement effect. In that study, however, we only explored the vowels [ɑ], [o] and [i]. We found that the unrounded back-open vowel [ɑ] and the rounded back-mid vowel [o] were linked to the backward hand movements while the unrouned front-close vowel [i] was linked to the forward hand movements. Hence, these findings showed a connection between processes that plan front-back direction components for hand movements and for tongue movements required for articulation. However, we did not explore whether similar interaction exists in relation to the roundness dimension of articulation. Although our previous studies show that the rounded back-mid vowel [o] is associated with backward hand movements, it is still possible that a rounded vowel is associated with forward movement if the tongue backness is controlled. Indeed, the mouth-gesture accounts discussed above emphasize that rounded vowels would be linked to forward hand movements whereas unrounded vowels would be linked to backward hand movements. The present study was designed to investigate this question. In Experiment 1, the participants were required to pronounce either the rounded front-mid vowel [ø] or the unrounded front-mid vowel [e] while moving their hand forward or backward. In that block, the only dimension that differentiates between the articulations associated with the two vowels is lip rounding. In another block, the participants were required to pronounce either the rounded back-close vowel [u] or the unrounded front-close vowel [i] while moving their hand forward or backward. The [i]-[u] block was included to test further the view, provided by our previous studies showing the connection between the rounded back-mid vowel [o] and the backward movements (Vainio et al., 2015), that the primary factor in the effect is the front/back position of the tongue rather than the rounding/unrounding of the lips. Indeed, it is possible that we would observe an interaction between the vowel [ø] and the forward movements when the tongue movements are controlled (the [e]-[ø] block), but not between [u] and forward movements in the [i]-[u] block even though the vowel [u] is similarly rounded as the vowel [ø]. That outcome would suggest that lip rounding is associated with forward hand movements but when lip rounding is combined with tongue backing, the tongue component of the articulation eliminates the lip rounding influence on the effect because it is even more tightly connected to the processes responsible for preparing directional hand movements.

We originally interpreted the articulation-movement effect to reflect interaction in processes that prepare direction-related components of forward-backward hand movements and front-back tongue movements for articulatory gestures (Vainio et al., 2015). However, that can be premature interpretation as the previous study did not reliably distinguish between the roles of openness and frontness dimensions of articulation in the effect. That is, because the vowels [ɑ] and [o] are both more back *and* open vowels than [i]. Consequently, the present study aims to validate the original interpretation and investigate whether the effect can be observed when the vocal responses differ only with respect to forward-backward tongue movement in absence of differences in vowel openness or roundness. Hence, in Experiment 1, the participants were required to pronounce the rounded front-mid vowel [ø] or the rounded back-mid vowel [o] while moving their hand forward or backward. If the articulation-movement effect is indeed based on programming frontness/backness dimension of the vowel production, the vowel [ø] should be linked to forward movements and the vowel [o] to backward movements.

Hand movements and consonant production (Experiment 2)

In addition to showing the interaction between forward-backward hand movements and pronunciation of different vowels, our previous study (Vainio et al., 2015) also revealed that the same effect can be observed in relation to pronunciation of consonants [t] vs. [k]. The apical consonant [t] was linked to relatively fast forward hand movements whereas the dorsal consonant [k] was linked to relatively fast backward hand movements. Given that [t] can be assumed to require more forward-directed tongue movement than [k], this finding was in line with the view that the articulation-movement effect reflects articulatory processes that push and pull the tongue for articulation. However, instead of reflecting interaction in programming movement direction of hand and tongue, it is also possible that this effect with the consonants is based on the semantic compatibility between the movement direction of the hand and the pronounced speech unit. In the study, the consonants were coupled with the vowel [e] (i.e., [te] and [ke]) because it is difficult to pronounce them alone. However, in Finnish the consonant-vowel (CV) syllable [te] means plural 'you', and consequently it might be assumed to be semantically associated with the movements that point away from the body. In line with this explanation, it has been shown, at least with Italian-speaking subjects, that reading the word 'LA' ('there') facilitates hand movements that are directed away from the body while reading the word 'QUA' ('here') facilitates hand movements that are directed toward the body (Chieffi, Secchi, & Gentilucci, 2009). Moreover, the fact that we did not replicate the articulation-movement effect with the syllables [te] and [ke] when the participants were native Czech speakers (Tiainen et al., 2017), although the effect was replicated with vowels in the same Czech speakers, supports this explanation because the syllable [te] does not have the same semantic meaning in Czech as in Finnish. Therefore, Experiment 2 investigates whether the interaction effect between pronounced speech units and horizontal hand movements can be observed in relation to pronouncing coronal and dorsal consonants by removing semantic content from the pronounced CV syllables.

We have also discovered the articulation-grip effect, similar to the articulation-movement effect, showing that certain articulatory gestures are associated with the precision (used to grasp small objects by pinching the object between the tips of the index finger and thumb) or power grip (used to grasp large objects by flexing fingers toward the palm) responses (Vainio et al., 2013). In this effect, participants were required to perform either precision or power grip response while they pronounced, for instance, the meaningless syllable [te] or [ke]. The consonant [t] was associated with relatively rapid precision grip responses whereas [k] was associated with the power grip responses. It was speculated that the articulatory gesture produced by bringing the tip of the tongue into contact with the alveolar ridge and the teeth (coronal consonant [t]) is associated with the precision grip because it provides an articulatory pinching gesture; the precision grip is analogously produced using tips of the index finger and the thumb. In contrast, the articulatory gesture for the voiceless velar stop consonant [k], produced by moving the back of the tongue against the velum, might be viewed as an articulatory counterpart for the power grip action because the power grip is produced by moving intermediate and proximal components of all fingers –rather than finger tips– against the palm base, i.e., both actions use the base part of the effector.

However, given that the syllable [te] has above mentioned semantic connotations, similarly to the articulation-movement effect, it is possible that the articulation-grip effect is based on semantic processes related to the used syllables. Therefore, Experiment 2 explores whether pronunciation of coronal vs. dorsal consonants can produce the articulation-grip as well as articulation-movement effect when semantic content is removed from the pronounced syllables. That is, the participants are required perform either the precision vs. power grip response or forward vs. backward hand movement, and to pronounce the syllables [tø] vs. [kø] (Experiment 2a) or the syllables [re] vs. [ke] (Experiment 2b). The consonant [r] was assumed to provide equally good match with the precision grip as the consonant [t] because Finnish [r] is a voiced alveolar trill consonant; That is, both of them are produced by fronting the tip of the tongue. The consonant [k] was assumed to provide good match with the power grip for the reasons discussed above. Finally, it is important to notice that all the syllables were entirely meaningless for Finnish participants.

Experiment 1

Methods

Participants and ethical review: Twenty naïve volunteers participated in the study (19–55 years of age; mean age = 26.3 years; 3 males; 1 left-handed). All participants were native speakers of Finnish and had normal or corrected-to-normal vision. Written informed consent was obtained from all participants. The study was approved by the Ethical Committee of the Institute of Behavioural Sciences at the University of Helsinki and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, stimuli and procedure: Each participant sat in a dimly lit room with his or her head 70 cm in front of a 19-in. CRT monitor (screen refresh rate: 85 Hz; screen resolution:

1280 × 1024). The head-mounted microphone was adjusted close to the participant's mouth for recording vocal responses. The manual response device was a joystick that was located between the participant and the monitor, 30 cm away from the monitor. The joystick was attached steadily onto the response table. The joystick was vertically positioned so that the up-most part of the stick was at the level of the low-most part of the monitor. It was positioned horizontally at the center of the monitor. Stimuli consisted of different target vowels [Block 1: i & u; Block 2: ö & e; Block 3: ö & o; (⟨ö⟩ is pronounced as [ø])] that were presented at the center of the monitor over a light-grey background color. The vowels were written in Consolas font (lowercase; bold; font size: 90). The vowels were presented in randomized order within all experimental blocks. In addition, the order of blocks was balanced between the participants. In total, Experiment 1 consisted of 360 trials [30 (repetition) × 3 (block) × 2 (vowel) × 2 (direction)].

All Participant held the joystick in their right hand. At the beginning of each trial, a black fixation cross was presented for 400 ms at the center of the screen. Then, a blank screen was displayed for 500 ms. After that the target vowel was presented at the center of the screen and remained in view for 1500 ms or until a response was made. Finally, a blank screen was displayed for 500 ms. The target was presented either in green or in blue color. The participants' task was to pull or push the joystick until the end of the full motion range of the joystick (4.5 cm forward or backward) as fast and accurately as possible according to the color of the stimulus. Half of the participants responded to the green by pulling the joystick, and other half responded to the blue by pulling the joystick. The response directions were marked with corresponding color tapes on the response device. The joystick was always returned to the central starting position after the response. The mechanisms of the joystick provided a minor force that pulled the stick back toward the starting position. All stimuli were displayed on a gray background. Erroneous manual responses were immediately followed by a short 'beep' tone.

Experiment 1 consisted of three blocks that were separated by a short break. The order of the blocks was balanced between participants. The participants were instructed to continue from the break by pressing with their left hand the space bar of a keyboard that was positioned between the joystick and the monitor. There was forced practice session of eight trials before the each block. In those sessions, the participants were able to shortly practice the task with a new set of vowels. In addition, both experiments began with practice trials. Each participant was given as much practice as it took to perform the task fluently. That is, the actual experiment was not started before the participant demonstrated in the practice session that he/she continuously produced the vocal and manual reactions relatively rapidly and at the same time.

In addition to the manual response, the participants were instructed to pronounce the presented vowel as fast as possible. It was emphasized that the vowel should be uttered in natural talking voice at the same time with the manual response. The recording levels of the vocal responses were calibrated individually for each subject at the beginning of the experiment. Stimulus presentation and sound recording were done with Presentation software. Vocal reaction times were measured for 1500 ms from the onset of the target object to the onset of the vocalization. The onset point of the vocalizations was searched and marked manually and separately for each trial.

Onsets of the vocalizations were located individually for each trial using Praat (v. 5.3.49) (http://www.praat.org). Manual reaction times (RTs) were measured from the onset of the target stimulus to the point of the joystick movement in which the joystick reached 100 % from the motion range of the joystick to the forward or backward direction. The response was registered as an 'error' if joystick movement exceeded 100 % of the motion range but was moved into the wrong direction. In addition, the response was registered as 'no-response' if the joystick movement did not reach 100 % of the motion range (see Fig 1).

---Figure 1 about here---

Results and discussion

In total, 8.2 % of the raw data were discarded from the analysis of *manual reaction times* including 0.9 % of trials containing vocal errors, 2.5 % of trials containing manual errors, and 4.8 % of trials in which the manual RTs were more than two standard deviations from a participant's overall mean. In total, 8.3 % of the raw data were discarded from the analysis of *vocal reaction times* including vocal and manual errors, and 4.9 % of trials in which the vocal RTs were more than two standard deviations from a participant's overall mean. Condition means for the remaining data were subjected to a repeated measures ANOVA with the within-participants factors of Block ([i]-[u], [ø]-[e] or [ø]-[o]), Vowel ([i]/[ø]/[ø] or [u]/[e]/[o]), and Manual response (push or pull, i.e., forward or backward, respectively). Post hoc comparisons were performed by means of *t* tests applying a Bonferroni correction when appropriate. A partial eta-squared statistic served as effect size estimate.

The analysis of vocal reaction times revealed a significant main effect of Vowel. Those vowels were pronounced faster that were hypothesized to be congruent with the backward movement ([u]/[e]/[o]) (M = 607 ms) rather than forward movement ([i]/[ø]/[ø]) (M = 613 ms), $F(1,19) = 8.17$, $MSE = 2848.48$, $p = .010$, $n_p^2 = .301$. The analysis also revealed a significant interaction between Vowel and Manual response [$F(1,19) = 65.54$, $MSE = 22173.71$, $p < .001$, $n_p^2 = .775$], and between Block, Vowel and manual response, $F(2,38) = 3.37$, MSE = 3104.87, $p = .045$, $n_p^2 = .151$. This interaction is presented in Fig 2. The pairwise comparisons test showed that, in Block 1, [i] responses were produced faster when the participants were simultaneously pushing the joystick (*M* = 587 ms) rather than pulling it (*M* = 622 ms) ($p < .001$). In contrast, [u] responses were produced faster when the participants were pulling the joystick (M = 579 ms) rather than pushing it (M = 606 ms) (p = .001). Similarly, in Block 3, [ø] responses were made faster when the participants were pushing the joystick (M = 610 ms) rather than pulling it (M = 636 ms) (p < .001).In contrast, [o] responses were produced faster when the participants were pulling the joystick (M = 614 ms) rather than pushing it (M = 629 ms) (p = .046). In Block 2, the interaction was not significant ([ø]-push: M = 611 ms; [ø]-pull: M = 614 ms, p = .719; ([e]-push: M = 609 ms; [e]-pull: M = 601 ms, p = .243). The errors were not analyzed for vocal responses because the participants produced so few (M = 0.9%) vocal errors.

The analysis of manual reaction times revealed a significant main effect of Vowel. The vowels that were hypothesized to be congruent with the backward movement ([u]/[e]/[o]) was pronounced faster (M = 586 ms) than the vowels that were hypothesized to be congruent with the forward movement ([i]/[ø]/[ø]) (M = 595 ms), $F(1,19) = 6.21$, $MSE = 4025.70$, $p = .022$, $n_p^2 = .246$. In addition, the pull responses were performed faster (M = 584 ms) than the push responses (M = 597 ms), $F(1,19) = 6.14$, $MSE = 9595.54$, $p = .023$, $n_p^2 = .244$. More importantly, the interaction between Vowel and Manual response was significant, $F(1,19) = 104.11$, $MSE = 30974,722$, $p < .001$, $n_p^2 = .846$. This interaction is presented in Fig 3. The push responses were performed faster when the vowel was [i] (*M* = 573 ms) rather than [u] (*M* = 594 ms) in Block 1, and [ø] (*M* = 594 ms) rather than [o] (*M* = 618 ms) in Block 3. However, this effect was statistically significant only in Block 3 (Block 1: p = .070; Block 3: p < .001). In Block 2, the push responses were not modulated by Vowel ([ø]: M = 601 ms; [e]: M = 600 ms, p = .918). In contrast, the pull responses were made faster what the vowel was [u] (m = 553 ms) rather than [i] (M = 596 ms) ($p < .001$) in Block 1, [e] (m = 574 ms) rather than [ø] (M = 594 ms) ($p = .013$) in Block 2, and [o] (m = 580 ms) rather than [ø] (M = 609 ms) ($p < .001$) in Block 3. The three-way interaction between Block, Vowel and Manual response was not significant, $F(1.439,27.342) = 2.52$, $MSE = 3606.09$, $p = .113$, $n_p^2 = .117$. The errors were not analyzed for manual responses because the participants produced so few (M = 2.5%) manual errors (12 participants produced under 10 errors). However, the mean error rates for manual responses are presented in Fig 3.

These results support our hypothesis that front vowels ([i] and [ø]) are associated with forward hand movements while back vowels ([u] and [o]) are associated with backward hand movements. These associations were observed in vocal and manual reaction times. However, there was no effect when the vowel production required lip rounding ([ø]) vs. spreading ([e]).

---Figure 2 about here---

---Figure 3 about here---

Experiments 2a & 2b

Methods

Participants and ethical review: In Experiment 2a, nineteen naïve volunteers participated in the study (20–37 years of age; mean age = 25.1 years; 3 males). In Experiment 2b, nineteen naïve volunteers participated in the study (19–39 years of age; mean age = 26.7 years; 2 males; 2 left-handed). All participants were native speakers of Finnish and had normal or corrected-to-normal vision. Written informed consent was obtained from all participants. The study was approved by the Ethical Committee of the Institute of Behavioural Sciences at the University of Helsinki and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus, stimuli and procedure: The stimuli and procedure were similar to Experiment 1 with the following exceptions: Firstly, in this study, the stimuli consisted of CV syllables in which the consonant content was manipulated [Exp. 2a: tö & kö; Exp 2b: re & ke (ö is pronounced as [ø])]. Secondly, Experiments 2a and 2b consisted of two blocks. The order of the blocks was balanced between participants. Similarly to Experiment 1, in the one block, the participants performed joystick responses. The response procedure in that block was identical to the procedure in Experiment 1. In the other block the participants performed grip responses. In the grip block, there were two response devices, each equipped with an inlaid microswitch: the precision grip device (1 × 1 × 0.7 cm) and the power grip device (10 cm long, 3 cm diameter). As the switches were depressed in each device, there was noticeable tactile feedback. The participant held these devices in their right hand that was resting on the table while holding the devices. Similarly to the joystick block, the participants' task was to respond as fast and accurately as possible according to the color of the stimulus. Half of the participants responded to the green with the precision grip and other half responded to the blue with the precision grip. The devices were marked with corresponding color tapes. The trial structure was identical in the blocks 1 and 2 of Experiment 2 as in the Experiment 1. In total, Experiment 2a consisted of 240 trials [30 (repetition) × 2 (block) × 2 (consonant) × 2 (response)] and Experiment 2b consisted of 240 trials [30 (repetition) × 2 (block) × 2 (consonant) × 2 (response)].

Results and discussion

Experiment 2a

Regarding the responses in the joystick block, in total, 6.0 % of the raw data were discarded from the analysis of *manual reaction times* including 0.5 % of trials containing vocal errors, 1.0 % of trials containing manual errors, and 4.5 % of trials in which the manual RTs were more than two standard deviations from a participant's overall mean. In total, 6.1 % of the raw data were discarded from the analysis of *vocal reaction times* including vocal and manual errors, and 4.6 % of trials in which the vocal RTs were more than two standard deviations from a participant's overall mean. Regarding the responses in the grip block, in total, 10.4 % of the raw data were discarded from the analysis of *manual reaction times* including 2.0 % of trials containing vocal errors, 4.1 % of trials containing manual errors, and 4.3 % of trials in which the manual RTs were more than two standard deviations from a participant's overall mean. In total, 10.7 % of the raw data were discarded from the analysis of *vocal reaction times* including vocal and manual errors, and 4.6 % of trials in which the vocal RTs were more than two standard deviations from a participant's overall mean.

The blocks were analyzed separately. Condition means for the remaining data were subjected to a repeated measures ANOVA with the within-participants factors Consonant ([tø] or & [kø]) and Manual response (Joystick block: push or pull; Grip block: precision or power). Post hoc comparisons were performed by means of *t* tests applying a Bonferroni correction when appropriate. A partial eta-squared statistic served as effect size estimate.

The joystick block: The analysis of vocal responses, did not reveal any significant main effects. In addition, the interaction between Consonant and Manual response was not observed, $F(1,18) = 0.82$, $MSE = 1550.93$, $p = .377$, $n_p^2 = .044$. This interaction is presented in Fig 4. Similarly, regarding manual responses, the main effects and the interaction between Consonant and Manual response [$F(1,18) = 1.48$, $MSE = 1907.49$, $p = .239$, $n_p^2 = .076$] were not significant. This interaction is presented in Fig 4. The errors were not analyzed for manual or vocal responses because the participants produced so few errors.

The grip block: The analysis of vocal reaction times revealed a significant interaction between Consonant and Manual response [$F(1,18) = 23.41$, $MSE = 63718.69$, $p < .001$, $n_p^2 = .565$]. This interaction is presented in Fig 5. The pairwise comparisons test showed that vocal responses were produced faster when the participants pronounced the consonant [t] and simultaneously performed the precision grip (M = 587 ms) rather than the power grip (M = 653 ms) ($p < .001$). In contrast, vocal responses were produced faster when the participants pronounced the consonant [k] and simultaneously performed the power grip (M = 585 ms) rather than the precision grip (M = 635 ms) (p = .009). The errors were not analyzed for vocal responses because the participants produced so few vocal errors.

The analysis of manual reaction times revealed a significant main effect of Manual response, $F(1,18) = 6.77$, $MSE = 11061.36$, $p = .018$, $n_p^2 = .273$. The precision grip responses were performed faster (M = 531 ms) than the power grip responses (M = 555 ms). In addition, the interaction between Consonant and manual response was significant, $F(1,18) = 36.27$, $MSE = 70319.98$, $p < .001$, $n_p^2 = .668$. This interaction is presented in Fig 5. The pairwise comparisons test showed that the precision grip responses were performed faster when the consonant was [t] (M = 501 ms) rather than [k] (M = 561 ms) (p < .001). In contrast, the power grip responses were made faster what the consonant was [k] (m = 524 ms) rather than [t] (M = 586 ms) (p < .001).

An analysis of percentage error rates of manual responses (notice that one participant did not make any errors) revealed a significant interaction between Vowel and Manual response, $F(1,17) = 23.66$, $MSE = 868.06$, $p < .001$, $n_p^2 = .582$. The pairwise comparisons test showed that the accuracy of manual responses was improved by the compatibility between the grip type and the consonant. The participants made fewer errors with the precision grip when the consonant was [t] (M = 0.6%) rather than [k] (M = 7.8%) (p < .001). In contrast, they made fewer errors with the power grip when the consonant was [k] (M = 1.1%) rather than [t] (M = 7.8%) (p = .002). The mean error rates for manual responses are however presented in Fig 5.

The results of Experiment 2a revealed that when the participants pronounced the consonants [t] or [k] and simultaneously moved their hand forward or backward, there was no interaction between the pronounced consonant and the movement direction. However, a clear and statistically significant interaction effect was observed between the pronounced consonant and the performed grip type. Vocal and manual responses were performed faster when the consonant was [t] and the manual response was the precision grip. Similarly, responses were performed faster when the consonant was [k] and the manual response was the power grip.

Experiment 2b

Regarding the responses in the joystick block, in total, 5.7 % of the raw data were discarded from the analysis of *manual reaction times* including 0.2 % of trials containing vocal errors, 1.2 % of trials containing manual errors, and 4.3 % of trials in which the manual RTs were more than two standard deviations from a participant’s overall mean. In total, 5.7 % of the raw data were discarded from the analysis of *vocal reaction times* including vocal and manual errors, and 4.3 % of trials in which the vocal RTs were more than two standard deviations from a participant’s overall mean. Regarding the responses in the grip block, in total, 7.4 % of the raw data were discarded from the analysis of *manual reaction times* including 0.9 % of trials containing vocal errors, 1.9 % of trials containing manual errors, and 4.6 % of trials in which the manual RTs were more than two standard deviations from a participant’s overall mean. In total, 7.4 % of the raw data were discarded from the analysis of *vocal reaction times* including vocal and manual errors, and 4.6 % of trials in which the vocal RTs were more than two standard deviations from a participant’s overall mean.

The blocks were analyzed separately. Condition means for the remaining data were subjected to a repeated measures ANOVA with the within-participants factors Consonant ([re] or & [ke]) and Manual response (Joystick block: push or pull; Grip block: precision or power). Post hoc comparisons were performed by means of *t* tests applying a Bonferroni correction when appropriate. A partial eta-squared statistic served as effect size estimate.

The joystick block: The analysis of vocal reaction times revealed a significant main effect of Consonant, $F(1,18) = 28.03$, $MSE = 31372.37$, $p < .001$, $n_p^2 = .609$. Vocal responses were made faster when the consonant was [r] (M = 568 ms) rather than [k] (M = 608 ms). The interaction between Consonant and manual response was not significant, $F(1,18) = 1.95$, $MSE = 2088.33$, $p = .180$, $n_p^2 = .098$. This interaction is presented in Fig 4. The analysis of manual reaction times revealed significant main effect of Manual response, $F(1,18) = 11.25$, $MSE = 4776.46$, $p = .004$, $n_p^2 = .385$. The pull responses were made faster (M = 595 ms) than the push responses (M = 611 ms). The interaction between Consonant and manual response was not significant, $F(1,18) = 1.36$, $MSE = 1325.36$, $p = .259$, $n_p^2 = .070$. This interaction is presented in Fig 4. The errors were not analyzed for manual or vocal responses because the participants produced so few errors.

The grip block: The analysis of vocal reaction times revealed a significant main effect of Consonant, $F(1,18) = 41.58$, $MSE = 24750.91$, $p < .001$, $n_p^2 = .698$. Vocal responses were made faster when the consonant was [r] (M = 574 ms) rather than [k] (M = 610 ms). In addition, a significant interaction between Consonant and Manual response was observed [$F(1,18) = 27.75$, $MSE = 47995.66$, $p < .001$, $n_p^2 = .607$]. This interaction is presented in Fig 5. The pairwise comparisons test showed that vocal responses were produced faster when the participants pronounced the consonant [r] and simultaneously performed the precision grip ($M = 550$ ms) rather than the power grip ($M = 598$ ms) ($p < .001$). In contrast, vocal responses were produced faster when the participants pronounced the consonant [k] and simultaneously performed the power grip (M = 583 ms) rather than the precision grip (M = 636 ms) (p < .001). The errors were not analyzed for vocal responses because the participants produced so few vocal errors.

The analysis of manual reaction times revealed a significant interaction between Consonant and Manual response, $F(1,18) = 19.20$, $MSE = 46008.66$, $p < .001$, $n_p^2 = .516$. This interaction is presented in Fig. 3b. The pairwise comparisons test showed that the precision grip responses were performed faster when the consonant was [r] (M = 528 ms) rather than [k] (M = 578 ms) (p = .001). In contrast, the power grip responses were made faster what the consonant was [k] (m = 526 ms) rather than [r] (M = 574 ms) (p < .001). The errors were not analyzed for manual responses because the participants produced so few vocal errors. The mean error rates for manual responses are however presented in Fig 5.

Similarly to Experiment 2a, the articulation-movement effect was not observed in relation to coronal [r] vs. dorsal [k] consonants. However, again similarly to the results of Experiment 2a, the coronal consonant was associated with precision grip responses and the dorsal consonant was associated with power grip responses.

---Figure 4 about here---

---Figure 5 about here---

General Discussion

The study validated our previous view (Vainio et al., 2015) that that the articulation-movement effect is linked to programming frontness-backness dimension of vowel production. In Experiment 1, when the participants were required to pronounce either the rounded front-mid vowel [ø] or the rounded back-mid vowel [o], the front vowel [ø] was associated with relatively rapid forward hand movements while the back vowel [o] was associated with relatively rapid backward hand movements. Similarly, in the [i]-[u] block of Experiment 1, the unrounded front-close vowel [i] was associated with forward hand movements while the rounded back-close vowel [u] was associated with backward hand movements. These findings support the view that there is interaction in programming direction of push and pull movements of a tongue and hand similarly to interaction in programming direction of simultaneous movements of two hands (Serrien et al., 1999). This finding expands the models of common direction coding (Cohen & Andersen, 2002) ─that have previously included movements of hands, head, eyes and legs─ to cover also direction coding processes related to tongue movements. That is, direction coordinates for articulatory tongue movements are not programmed in isolation but share processes, at least, with planning of hand movements. This finding also underlines that hand and mouth actions do not only interact at the level of finger shaping for hand opening/closing (Gentilucci et al., 2001) or performing a specific hand grip (Vainio et al., 2013) but that hand and mouth actions also interact at the level of programming direction for push and pull movements of the tongue and hand.

The articulation-movement effect does not appear to operate in relation to programming lip rounding for vowel production. In Experiment 1, production of rounded ([ø]) vs. unrounded ([e]) front-mid vowels was not influenced by forward-backward hand movements. This supports the view that the effect operates in relation to the horizontal tongue movements rather than lip rounding. However, the original mouth-gesture theories (Wallace, 1881; Ramachandran & Hubbard, 2001) have emphasized a potential interaction between deictic hand gestures and the speech processes related to rounding component of the articulatory gesture so that rounded vowels would be linked to the outward pointing hand movements and unrounded vowels would be linked to the inward pointing hand movements. The present finding suggests that the articulation-movement effect does not operate at the level of gestural iconicity: commonalities in the iconicity between articulation and horizontal hand movement –that is, between lip protrusion gesture and outward hand movement– do not trigger the effect. Rather, the effect is bound to programming tongue fronting-backing movements that mostly occur invisibly inside the mouth. However, the ideas of the mouth-gesture theories that associate lip protrusion movements with outward pointing gestures are mostly based on theoretical speculation rather than hard data. In the light of the current evidence, we suggest that in order to validate the mouth-gesture theories assuming that a vocabulary might have been shaped by the tendencies to mimic with articulators, for example, the outward-inward hand pointing actions (e.g., Wallace, 1881; Ramachandran & Hubbard, 2001), researches should focus on exploring statistical frequencies of front and back vowels in the words that are semantically pointing outwards or inwards of one's own body instead of focusing on frequencies of rounded-unrounded vowels in these words. In line with this view, Wichmann, Holman and Brown (2010) have reported the cross-linguistic study that did not show evidence for associating the vowel roundness with any deictic words. However, as our findings would predict, the back vowel [ɑ] was relatively frequently included in the pronoun that points to the speaker ('me') whereas the front vowel [i] was more frequently included in the pronoun that points to the hearer ('you').

In most cases, the articulation-movement effect as well as the articulation-grip effect were observed in both vocal and manual reaction times. Regardless, it is possible that the effect ultimately operates only in relation to one response type (e.g., manual responses), and the effect related to the other is just a byproduct of the congruency effect related to the first one, driven by the task which requires that participants intentionally synchronize the two responses. However, we have previously shown that at least the articulation-grip effect can be observed in vocal responses even when the grip is planned beforehand but not actually executed (Tiainen et al., 2016). In addition, this effect is observed in manual reaction times when the participant reads silently a congruent or incongruent syllable prior executing precision and power grip responses according to the color of a visual target, and when the participant hears these syllables and performs grip responses according to the pitch of the auditory stimuli (Vainio, Tiainen, Tiippana, & Vainio, 2014). This suggests that at least the articulation-grip effect related to manual and vocal responses reflects bi-directional influences between vocal and manual processes rather than task demands requiring synchronization between two responses. However, it cannot be concluded indisputably whether the articulation-movement effect observed ultimately reflects the influence of vowel production on manual responses, the influence of manual responses on vowel production or whether the influence is bi-directional.

The congruency effects related to the consonant articulation

The results of Experiment 2 showed that the articulation-movement effect is exclusively related to the processes that program articulatory gestures for vowel production, and the articulation-movement effect that was previously observed with consonants (Vainio et al., 2015) was most likely based on semantic connotations of the used syllables. The articulation-movement effect was entirely missing when the semantic meaning was removed from the pronounced CV syllables. In sharp contrast, when these same CV syllables were pronounced simultaneously with the precision and power grip responses, the consonants [t] and [r] were linked to precision grip responses while the consonant [k] was linked to power grip responses. These findings related to grip responses support the view that the articulation-grip effect, regarding consonant production, reflects processes that program tongue shape for producing coronal vs. dorsal stop closures (Vainio et al., 2013).

Why then coronal vs. dorsal consonants were not associated with the forward vs. backward hand movements, respectively, while front vs. back vowels were systematically associated with them? The coronal consonants are nonetheless produced by fronting the tongue tip toward high-front position while the dorsal consonants are produced by raising the tongue body toward the back of the velum. It should be emphasized that the tongue muscles involved in producing vowels are to some extent different from the muscles involved in producing tongue shape for consonants in general and the coronal-dorsal consonants in particular. It is assumed that tongue is moved in vowel production mostly by extrinsic tongue muscles such as the genioglossus and the styloglossus (Sauerland & Mitchell, 1975). In short, these extrinsic muscles are responsible for positioning the tongue body in the oral cavity for vowel production; that is, to move the tongue by fronting or backing (i.e., movement in the horizontal axis) and raising or lowering (i.e., movement in the vertical axis) the tongue body in order to provide a specific horizontal-vertical movement of the tongue body for each vowel. In contrast, the intrinsic tongue muscles such as the superior longitudinal muscle and the inferior longitudinal muscle are mostly responsible for shaping the tongue for producing the consonants such as [t] and [k] (Perkell, 1969; Sauerland & Mitchell, 1975). It has been also found that consonants and vowels are represented in distinct regions in the speech production system (Bouchard, Mesgarani, Johnson, & Chang, 2013). In the light of these findings, we propose that the articulation-movement effect mostly operates in relation to controlling extrinsic tongue muscles for producing horizontal-vertical movements of the tongue body in the generation of vowels. In other words, we assume that the forward-backward movements of proximal arm joints and horizontal-vertical movements of the tongue body interact in planning their direction-related components.

In line with the view presented above, Öhman (1966) has suggested that vowels and consonants are produced by three tongue articulators: body, apex and dorsum. The body is mostly responsible for producing movements of the tongue body for vowels, while the apex (related to tongue tip movements) and the dorsum (related to arching of the tongue body) are mostly involved in the generation of consonants. In the light of this account, we propose that the articulation-grip effect observed in relation to pronouncing the coronal vs. dorsal consonants is based on action planning processes that are to some extent shared by programming movements of fingers for producing manual grasping and movements of intrinsic articulators (i.e., apex and dorsum) for producing consonants. In more detail, we assume that there is an interaction between processes that plan apex-related tongue movements (e.g., fronting the tongue tip in order to produce coronal stop colure) and the precision grip, and between processes that plan dorsum-related tongue movement (e.g., arching the tongue body in order to produce dorsal stop closure) and the power grip. According to this speculative view, the articulation-grip effect is a consequence of programming simultaneously the articulatory and manual actions that share the same action planning processes; e.g., the apex-related tongue action for producing the consonant [t] and the finger (i.e., distal hand joints) movements utilizing the tips of the index finger and the thumb for producing the precision grasp. As another consequence of these between-effector interactions, a relative delay in responses can be observed when the two simultaneously programmed actions employ different action planning processes; e.g., dorsum-related tongue action for producing the consonant [k] and the finger movements for producing the precision grasp.

An alternative explanation for the articulation-grip effect, which is not mutually exclusive with the above discussed “apex-dorsum” hypothesis, assumes that the neural connection between a specific grip type and a tongue shape is based on motor planning and execution mechanisms that are required for producing tongue and hand actions at different levels of complexity. According to this view, the motor processes responsible for performing the precision grip, which is a relatively complex hand action (Marzke, 1997), are also utilized for performing the apex-related tongue actions. That is, because both of these actions require relatively fine motor coordination, accurate finger/tongue positioning and/or great control on muscle activity, at least compared to the power grip and the dorsum-related tongue movements. Indeed, increased co-contractions of antagonist muscles are required for performing hand actions with an increased accuracy (Gribble, Mullin, Cothros & Mattar, 2003). Complementary to this notion, it has been shown that precision grip has a larger cortical representation than power grip both in monkeys (Fluet, Baumann, & Scherberger, 2010; Rizzolatti, et al., 1988) and in humans (Pistohl, Schulze-Bonhage, Aertsen, Mehring, & Ball, 2012). In other words, it is possible that the precision grip is associated with the apex-related articulatory tongue movements because both actions demand relatively fine motor coordination while the power grip is associated with the dorsum-related tongue movements because both are based on more coarse motor coordination processes.

Conclusion

The present findings suggested that programming forward-backward hand movements and tongue movements for producing front-back vowels share the same direction codes in action planning. Vowel articulation and horizontal hand movements are produced relatively rapidly when the vowel production requires tongue fronting and the hand is moved forwards. Similarly, articulation and hand movements are produced relatively rapidly when the vowel production requires tongue backing and the hand is moved backwards. However, the action planning processes that program lip rounding/spreading were not connected to the processes that program horizontal hand movements.

The articulation-movement effect was exclusively linked to vowel production. When the articulation task required pronunciation of coronal (i.e., [t] or [r]) vs. dorsal (i.e., [k]) consonants, that similarly to the front-back vowels can be assumed to require relative tongue movements within the front-back axis, the effect was only observed in relation to the precision vs. power grip responses, while the effect was entirely missing in relation to horizontal hand movements.

In the light of the current findings, we propose that the horizontal movements of the tongue body for vowel production share the movement planning processes with the reach-related hand movements whereas the tongue articulators of the apex (related to tongue tip movements) and dorsum (related to arching of the tongue body) share the action planning processes with the precision and power grasp, respectively.

Acknowledgement and funding details

The research leading to these results has received funding from the Academy of Finland under grant agreement number 1265610.

References

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature, 495(7441),* 327-332.

Chieffi, S., Secchi, C., & Gentilucci, M. (2009). Deictic word and gesture production: Their interaction. *Behavioural brain research, 203(2),* 200-206.

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience, 3(7),* 553-562.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual review of neuroscience, 22(1),* 319-349.

Corfield, D.R., Murphy, K., Josephs, O., Fink, G.R., Frackowiak, R.S., Guz, A., Adams, L. & Turner, R. (1999). Cortical and subcortical control of tongue movement in humans: a functional neuroimaging study using fMRI. Journal of Applied Physiology, 86(5), 1468-1477.

Fluet, M. C., Baumann, M. A., & Scherberger, H. (2010). Context-specific grasp movement representation in macaque ventral premotor cortex. *Journal of Neuroscience, 30*, 15175-15184.

Gentilucci, M., Benuzzi, F., Gangitano, M. & Grimaldi, S. (2001). Grasping with hand and mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology, 86,* 1885–1699.

Gribble, P. L., Mullin, L. I., Cothros, N., & Mattar, A. (2003). Role of cocontraction in arm movement accuracy. *Journal of neurophysiology, 89*, 2396-2405.

Heuer, H. (1982). Binary choice reaction time as a criterion of motor equivalence. *Acta Psychologica, 50,* 35-47.

Hewes, G.W. (1973). Primate communication and the gestural origins of language. *Current Anthropology, 14,* 5–24.

Higginbotham, D.R., Isaak, M.I., & Domingue, J.M. (2008). The exaptation of manual dexterity for articulate speech: an electromyogram investigation. *Experimental Brain Research, 186*, 603–609.

Kawashima, R., Itoh, H., Ono, S., Satoh, K., Furumoto, S., Gotoh, R., Koyama, M., Yoshioka, S., Takahashi, T., Takahashi, K. & Yanagisawa, T. (1996). Changes in regional cerebral blood flow during self-paced arm and finger movements. A PET study. *Brain research, 716(1),* 141-148.

Marzke, M. W. (1997). Precision grips, hand morphology, and tools. *American Journal of Physical Anthropology, 102*, 91-110.

Paget, R. (1930). *Human speech.* New York: Harcourt Brace.

Perkell, J. S. (1996). Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling. *Journal of Phonetics, 24(1),* 3-22.

Petrides, M., & Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biology, 7(8),* e1000170.

Pistohl, T., Schulze-Bonhage, A., Aertsen, A., Mehring, C., & Ball, T. (2012). Decoding natural grasp types from human ECoG. *Neuroimage, 59*, 248-260.

Ramachandran, V.S., & Hubbard, E.M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of Consciousness Studies, 8,* 3-34.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research, 71,* 491–507.

Rosenbaum, D. A. (1980). Human movement initiation: Specification of arm, direction, and extent. *Journal of Experimental Psychology: General, 109*, 444-474.

Sauerland, E. K., & Mitchell, S. P. (1974). Electromyographic activity of intrinsic and extrinsic muscles of the human tongue. *Texas reports on biology and medicine, 33(3),* 444-455.

Serrien, D. J., Bogaerts, H., Suy, E., & Swinnen, S. P. (1999). The identification of coordination constraints across planes of motion. *Experimental Brain Research, 128(1),* 250-255.

Swinnen, S. P., Jardin, K., Meulenbroek, R., Dounskaia, N., & Hofkens-Van Den Brandt, M. (1997). Egocentric and allocentric constraints in the expression of patterns of interlimb coordination. *Journal of Cognitive Neuroscience, 9(3),* 348-377.

Tiainen, M., Tiippana, K., Vainio, M., Komeilipoor, N., & Vainio, L. (2017). Interaction in planning vocalizations and grasping. *The Quarterly Journal of Experimental Psychology, 70*, 1590-1602.

Tiainen, M., Lukavský, J., Tiippana, K., Vainio, M., Šimko, J., Felisberti, F., & Vainio, L. (2017). Systematic connections of grasping and horizontal hand movements with articulation in Czech speakers. *Frontiers in Psychology.* 5;8:516.

Vainio, L., Schulman, M., Tiippana, K., & Vainio, M. (2013). Effect of syllable articulation on precision and power grip performance. *PLoS ONE 8(1),* e53061.

Vainio, L. Tiainen, M., Tiippana, K., & Vainio, M. (2014). Shared processing of planning articulatory gestures and grasping. *Experimental Brain Research, 232,* 2359–2368.

Vainio, L., Tiainen, M., Tiippana, K., Komeilipoor, N., & Vainio, M. (2015). Interaction in planning movement direction for articulatory gestures and manual actions. *Experimental brain research, 233(10),* 2951-2959.

Wakelin, D.R. (1976). The role of the response in psychological refractoriness. *Acta Psychologica, 40,* 163-175.

Wallace, A.R. (1881) Review of Tylor's Anthropology. *Nature, 24,* 242-245.

Waters, G.S., & Fouts, R.S. (2002). Sympathetic mouth movements accompanying fine motor movements in chimpanzees (Pan troglodytes) with implications toward the evolution of language. *Neurological Research, 24*, 174–180.

Watanabe, J., Sugiura, M., Miura, N., Watanabe, Y., Maeda, Y., Matsue, Y., & Kawashima, R. (2004). The human parietal cortex is involved in spatial processing of tongue movement—an fMRI study. *Neuroimage, 21(4),* 1289-1299.

Wichmann, S., Holman, E. W., & Brown, C. H. (2010). Sound symbolism in basic vocabulary. *Entropy, 12(4),* 844-858.

Öhman, S. E. (1966). Coarticulation in VCV utterances: Spectrographic measurements. The *Journal of the Acoustical Society of America, 39(1),* 151-168.

Figure captions

Figure 1. Graphical presentation of the trial structure of Experiments 1 and 2. See text for details.

Figure 2. The mean vocal reaction times for Experiment 1 as a function of the vowel and the movement direction. Error bars depict the standard error of the mean. Asterisks indicate statistically significant differences (*** <. 001; **p < .01; *p < .05).

Figure 3. The mean manual reaction times for Experiment 1 as a function of the vowel and the movement direction. Error bars depict the standard error of the mean. Asterisks indicate statistically significant differences (*** <. 001; **p < .01; *p < .05). The means for percentage error rates of manual responses is presented below the corresponding reaction time histogram.

Figure 4. The mean vocal and manual reaction times for Experiment 2 as a function of the syllable and the movement direction. Error bars depict the standard error of the mean.

Figure 5. The mean vocal and manual reaction times for Experiment 2 as a function of the syllable and the grip type. Error bars depict the standard error of the mean. Asterisks indicate statistically significant differences (*** <. 001; **p < .01; *p < .05). The means for percentage error rates of manual responses is presented below the corresponding reaction time histogram of grip responses.

Figure 1

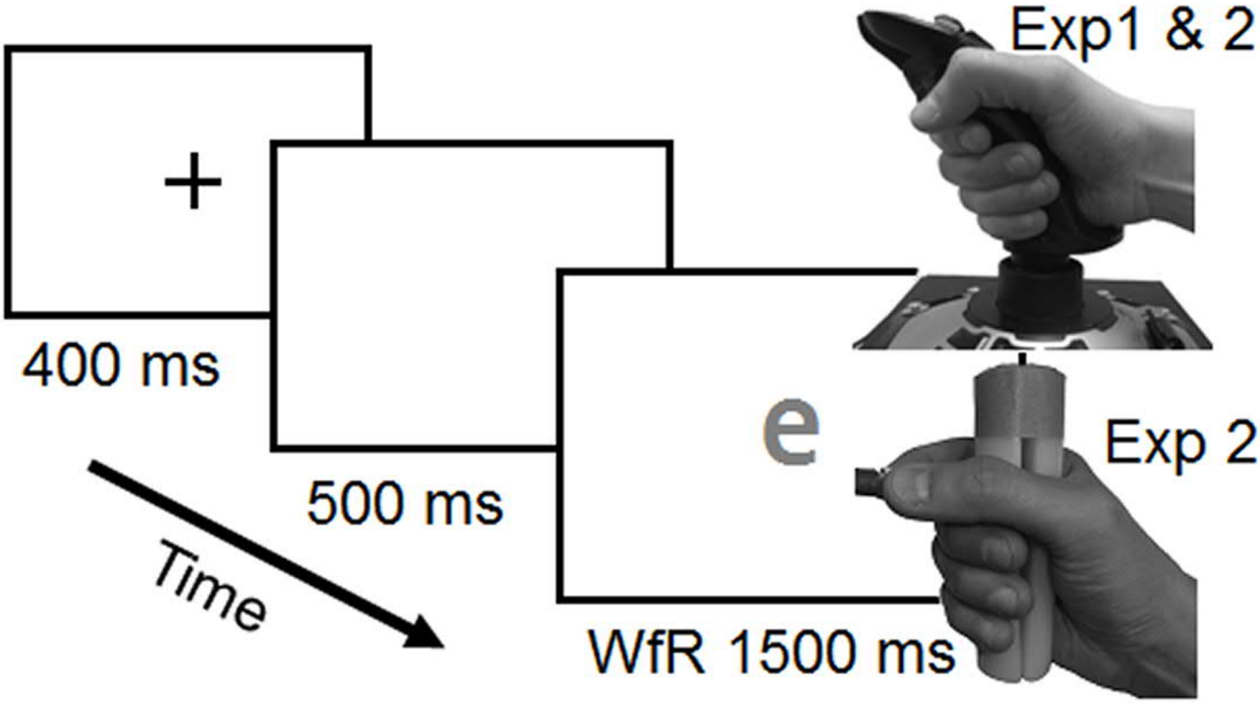


Figure 2

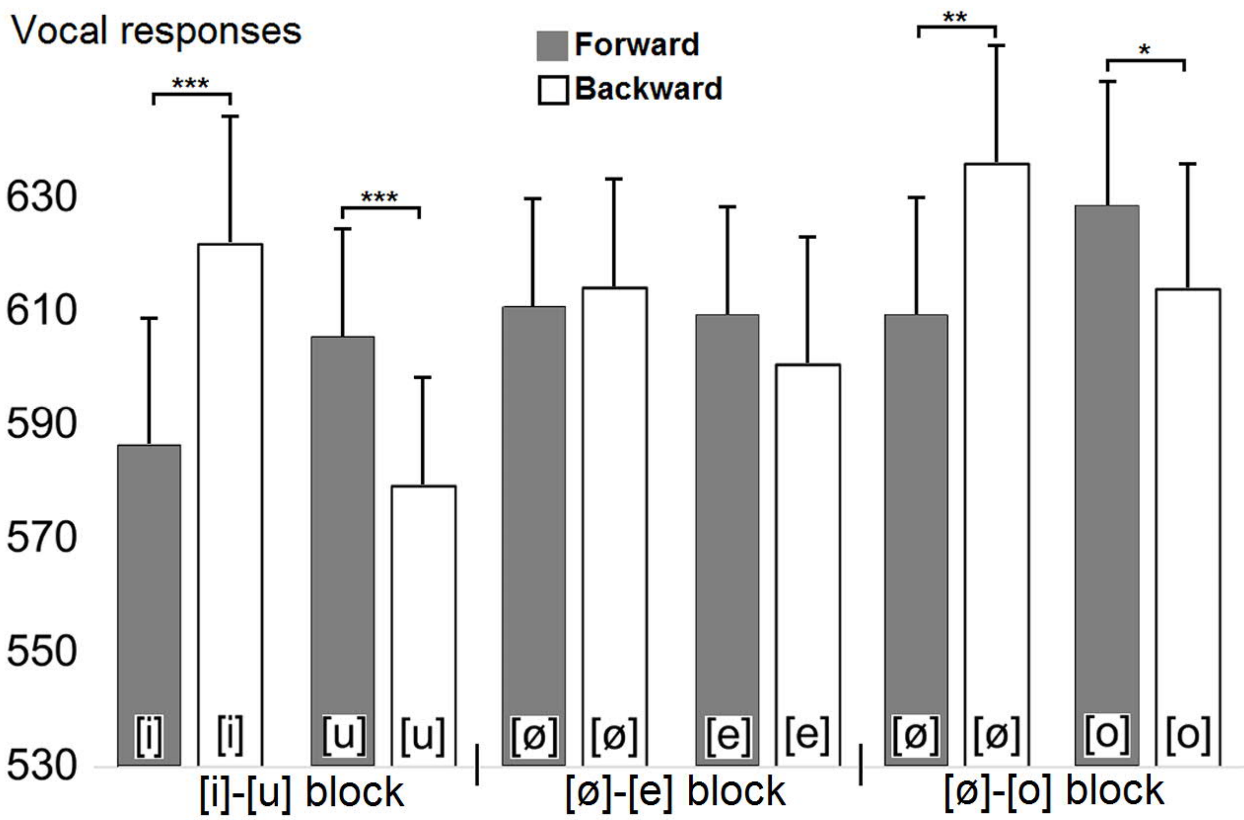


Figure 3

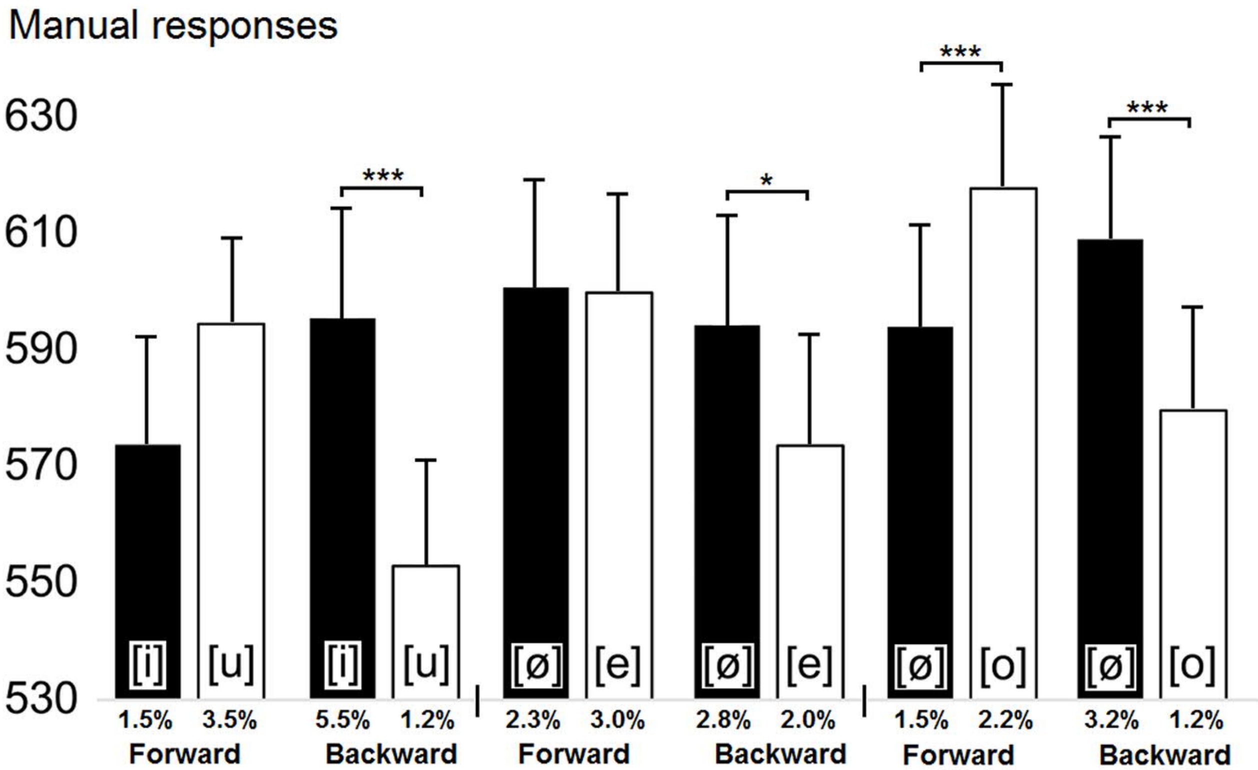


Figure 4

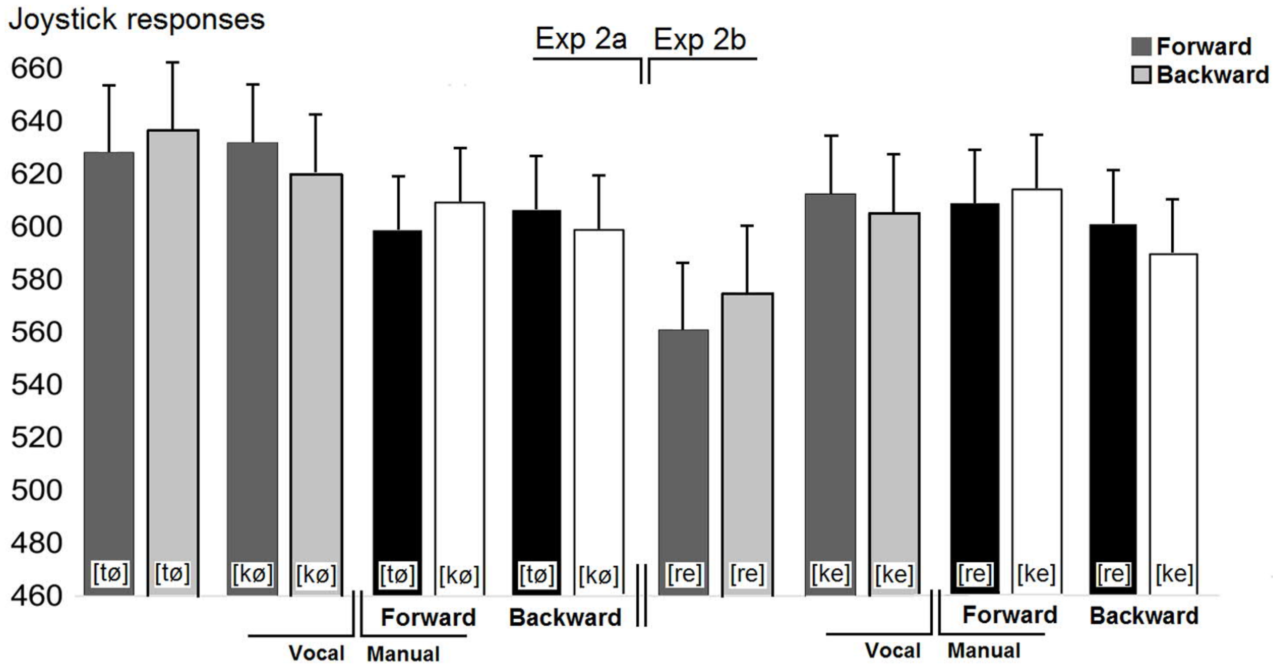
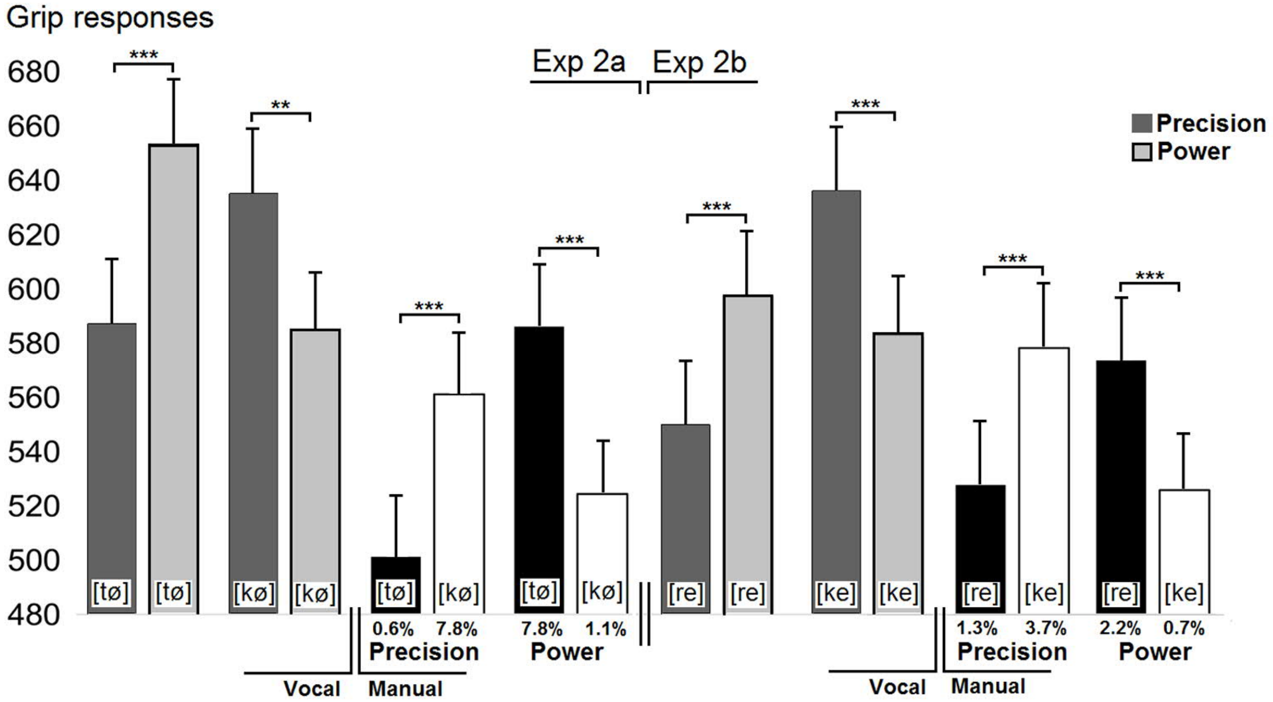


Figure 5


Grip responses
680
660
640
620
600
580
560
540
520
500
480
Exp 2a
Exp 2b
Precision
Power
***
**
***
***
***
***
***
***
[tø] [tø] [kø] [kø] [tø] [kø] [tø] [kø] [re] [re] [ke] [ke] [re] [ke] [re] [ke]
0.6% 7.8% 7.8% 1.1% 1.3% 3.7% 2.2% 0.7%
Precision
Power
Vocal
Manual
Precision
Power
Vocal
Manual